Ja Yun Choi1, Hyeonseok Jin[2], Kyungbaek Kim[2], Sola Cho[1], **Sehyun Cho*[1]**

[1]College of Nursing, Chonnam National University · Chonnam Research Institute of Nursing Science, Gwangju, 61469, Republic of Korea

[2]Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, 61186, Republic of Korea

Enhancing Deep Learning Models for Predicting Smoking Status Using Clinical Data in Patients with COPD

Background

Chronic obstructive pulmonary disease (COPD) is a major health concern, with smoking as its most critical modifiable risk factor. Advanced deep learning approaches integrating clinical, behavioral, and psychosocial factors are needed to improve predictive performance.

Purpose

This study aimed to develop and evaluate deep learning models to improve the prediction of persistent smoking in patients with COPD, using clinical data from a prospective survey that incorporates behavioral and psychosocial variables along with variables from our previous model based on a structured national dataset.

Methodology

Three deep learning models and one machine learning model were developed and evaluated using clinical, behavioral, and psychosocial variables from 350 patients with COPD, including 51 current smokers. Data preprocessing involved imputing missing values, transforming variables, and adjusting for class imbalance. Hyperparameter optimization was performed using the Optuna framework. Model performance was evaluated through repeated stratified K-fold cross-validation to ensure a narrower 95% confidence interval. The macro F1 score was used as the primary evaluation metric. Additionally, Shapley Additive Explanations (SHAP) were applied to assess feature importance and enhance model interpretability.

Results

The Residual Neural Network (ResNN) model achieved the highest performance, with a macro F1 score of .87 (95% CI: .83–.89), followed by FT Transformer (.87, CI: .80–.92), XGBoost (.80, CI: .77–.83), TabTransformer (.78, CI: .74–.81). SHAP analysis identified professional advice to quit, employment status, sputum symptoms persisting for more than three months, perceived stress level, health check-up experience, and health literacy as key predictors of persistent smoking, particularly in the ResNN model.

Conclusion

This model integrating behavioral and psychosocial data outperformed our previous model by better capturing complex smoking patterns while maintaining interpretability. These findings highlight the importance of multidimensional data for identifying high-risk patients and guiding targeted smoking cessation strategies in COPD care.

doi.org/10.51219/URForum.2025.Sehyun-Cho ISBN: 978-1-917892-34-6

Keywords
classification, deep learning, machine learning, chronic obstructive pulmonary disease, smoking

Biography
SEHYUN CHO received her B.S. degree in Nursing from Honam University, Gwangju, Republic of Korea. She is currently a Ph.D. candidate in the integrated M.S.-Ph.D. program at the College of Nursing, Chonnam National University. She worked as a clinical nurse at Chonnam National University Hospital for five years. Her research interests include chronic disease management, inhaler use behavior, patient education, and the application of artificial intelligence in healthcare.

doi.org/10.51219/URForum.2025.Sehyun-Cho ISBN: 978-1-917892-34-6